

Evaluation of Phonatory Behavior of German and French Speakers in Native and Non-native Speech

Manfred Pützer¹, Frank Zimmerer¹, Wolfgang Wokurek², Jeanin Jügler¹

¹ Computational Linguistics and Phonetics, Saarland University, Saarbrücken, Germany

² Institute for Natural Language Processing, University of Stuttgart, Stuttgart, Germany

[puetzer, zimmerer, juegler]@coli.uni-saarland.de, wolfgang.wokurek@ims.uni-stuttgart.de

Abstract

Phonatory behavior of German speakers (GS) and French speakers (FS) in native (L1) and non-native (L2) speech was instrumentally examined. Vowel productions of the two groups were analyzed using a parametrization of phonatory behavior and phonatory quality properties in the acoustic signal. The behavior of GS is characterized by more strained adduction of the vocal folds whereas FS show more incomplete glottal closure. Furthermore, GS change their phonatory behavior in the foreign language (=French) by adapting phonatory strategies of FS, whereas FS do not show this tendency. In addition, German beginners (BEG) and partly German advanced learners (ADV) are already orientated on production characteristics of the L2. French BEG however retain their phonatory behavior in L2 (=German) by showing less vocal fold adduction in comparison to their L1. French ADV show the opposite behavior. Finally, ADV of the two speaker groups generally show more strained behavior in L2 productions than BEG. The results provide evidence that GS and FS apply different laryngeal phonatory settings and that they altered their settings in L2 differently. Perceptual evaluation of voice quality of the speech material and a correlation analysis between acoustic and perceptual results are suggested for future research.

Index Terms: Phonatory behavior, language learning, German, French

1. Introduction

Learning a foreign language in adulthood is a difficult task. Reaching native-like skills with respect to the phonetic and phonological system of the L2 is especially difficult. One of the reasons for this challenge is connected to the phonetic and phonological system of the native language. It has been shown that the L1 system may severely interfere with the phonetic and phonological system of the L2, possibly leading to an audible foreign accent (e.g., among many more [1-3]). An audible accent can be based on incorrect productions at the segmental and the suprasegmental level. A well-known example are /l/ and /r/, which are allophones for Japanese native speakers and thus very difficult to learn in an L2 like English (e.g., [4, 5]).

Incomplete acquisition of prosodic features in an L2 has also been identified as a source of a perceptible foreign accent (e.g., [6-9]). This includes differences concerning pitch range, or the correct location of pitch changes which may be difficult to learn (e.g., [10-17]). One pattern that has been found to occur in productions by L2 learners is a suppressed pitch range in L2 compared to the pitch range in the L1 [17-20]. One possible

explanation for this suppression is that L2 learners are less confident about speaking the foreign language, or that they focus on getting the segmental pronunciation and the placement of stress correctly before expanding the pitch range as native speakers do.

Phonatory quality has also been identified as a possible source for foreign accent in L2 productions (e.g., [9, 21-23]). For instance, both Brazilian Portuguese learners of English as well as English learners of Brazilian Portuguese produced different Long-Term Average Spectrum (LTAS) measures (of voice quality) in their respective L2 compared to productions in L1 [23]. Similarly, Italian and German speakers produced different patterns in their respective L2 [22]. Explanations of why these differences occur focus on different aspects. One reason is connected to the differences between languages themselves, which can lead to interferences and challenges in language acquisition. As for pitch variation, another reason is based on insecurity when speaking in an L2. However, there is still a lack of research which investigates different aspects of phonatory quality in an L1-L2 setting [21]. Therefore, we investigate French native speakers learning German and German native speakers learning French and compare their L1 and L2 productions. We chose a parametrization of phonatory behavior and phonatory quality properties in the acoustic signal based on the procedure developed by Wokurek and Pützer [24-27]. The following *four* research questions concerning phonatory behavior of German speakers (GS) and French speakers (FS) in native and non-native speech will be answered by the present study:

Firstly, do phonatory behavior differences appear when the productions of GS producing vowels in German words (GW) are compared with those of FS producing vowels in French words (FW)?

Secondly, do phonatory behavior differences appear when the productions of GS producing vowels in GW are compared with their productions of vowels in FW? The same question is conversely applied to FS. Their vowel productions in FW are compared with their vowel productions in GW.

Thirdly, do phonatory behavior differences appear (a) for GS beginners (BEG) and (b) for advanced learners (ADV), when the productions of vowels in GW are compared with their productions of vowels in FW? The same question is conversely applied to FS concerning their vowel productions in GW.

Fourthly, do phonatory behavior differences appear for ADV of the two groups (GS and FS) when their vowel productions in L2 are compared with the vowel productions of BEG in L2?

2. Material and Methods

2.1. Speakers and speech material

Our investigation of phonatory behavior patterns for French speakers (FS) and German speakers (GS) is based on the IFCASL learner corpus [28, 29]. The corpus consists of read speech in both the native language and the foreign language of FS and GS. For the analysis, the productions of 81 adult speakers were analyzed. Of these, 44 were French native speakers learning German (22 male, 22 female), 22 FS (10 male and 12 female) were advanced learners (ADV – B2-C2 according to the Common European Framework of Languages (CEFR)). The other 22 FS (12 male and 10 female) were beginners (BEG – A1-B1 according to CEFR). 37 participants were German learners of French (18 male, 19 female), 17 ADV (8 male, 9 females) and 20 BEG (10 male, 10 female). The corpus consists of 120 sentences per speaker, which are either read aloud in each language (*read condition*, such as “*In Berlin zahlt man wenig Miete*”; ‘In Berlin, you pay little rent’), or they were read after listening to a native model (*repeat condition*, such as “*Elle habite dans un beau village en France*”; ‘She lives in a nice village in France’). For the native productions, all sentences were read without listening to a model speaker.

From these sentences, 20 words were selected (10 in German, 10 in French). Half of them occurred in sentences in the *read condition*, the other half occurred in the *repeat condition*. All of the words were mono- or disyllabic. The vowel of interest was always [a:, a] occurring in the prominent syllable of the word. The vowel of interest was extracted automatically with a PRAAT [30] script, based on an automatic alignment of the corpus [29].

2.2. Acoustic measurements

Phonatory behavior is parameterized in this study by the bandwidth of the first formant and by spectral gradient features of the source spectrum [24]. The original parameters were defined by Stevens and Hanson [31]. They notice that different aspects of phonatory quality dominate certain parts of the glottal excitation spectrum. Five gradient parameters try to capture the overall decaying shape of the source spectrum. One bandwidth parameter captures the damping of the first formant introduced by the incomplete closed glottis. Our extension differs in a more complete compensation of the vocal tract filter and in frequency normalizations. The more complete compensation of the vocal tract filter uses the first four formant frequencies F_1, \dots, F_4 and formant bandwidths B_1, \dots, B_4 from LPC estimation. All four formants will be compensated to closely approach the source spectrum.

First, a spectrum is calculated using a 25ms Hamming window, long enough to show the harmonic structure. Figure 1 shows a schematic harmonic speech spectrum. Denoted are the amplitudes and frequencies that are used for the calculation of phonatory quality parameters: the first harmonic or fundamental oscillation H_1 at F_{0P} , the second harmonic H_2 at $2F_{0P}$, and the harmonic next to each of the first four formants A_{1P}, \dots, A_{4P} at F_{1P}, \dots, F_{4P} .

Figure 1 shows a schematic harmonic speech spectrum. Denoted are the amplitudes and frequencies that are used for the calculation of phonatory quality parameters: the first harmonic or fundamental oscillation H_1 at F_{0P} , the second harmonic H_2 at $2F_{0P}$, and the harmonic next to each of the first four formants A_{1P}, \dots, A_{4P} at F_{1P}, \dots, F_{4P} .

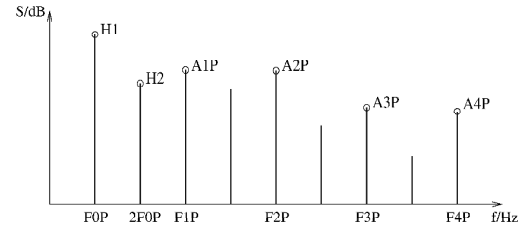


Figure 1: Schematic harmonic speech spectrum.

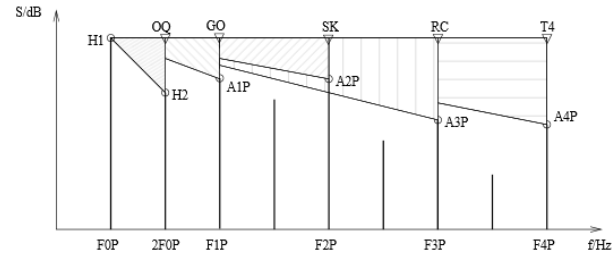


Figure 2: Schematic harmonic voice spectrum and phonatory behavior and quality parameter definitions.

Now the gain of the four formants is subtracted (in decibel amplitude scale) to estimate the source spectrum shown schematically in Figure 2.

Figure 2 shows the triangle slopes which are our phonatory behavior and quality parameters, respectively. We keep the parameter names of Stevens and Hanson [31]. OQ stands for open quotient and the spectral slope parameter is OQG with the appended G for gradient. The unit is decibels per barks because the perceptual bark scale is used. GO stands for glottal opening and defines the phonatory behavior parameter GOG. SK stands for skewness, RC for rate of closure. In our experience, these names do not correlate for every speaker to what they seem to address in the glottal waveform. We introduce a neutrally named parameter T4 (triangle at the fourth formant), and the phonatory quality parameter is T4G.

Besides these spectral source spectrum shape parameters OQG, GOG, SKG, RCG, and T4G, the bandwidth of the first formant is normalized by the frequency of the first formant and is called ‘incompleteness of closure’ (IC) = B_1/F_1 .

All normalizations were included as an attempt to reduce the influence of vowel quality on the phonatory behavior and quality parameters.

2.3. Statistical procedures

Data analysis was performed using SPSS version 22 for all tests. For each of the parameters as a dependent variable, ANOVAs and/or nonparametric tests (Mann-Whitney-U tests, depending on normal distribution) were carried out for the effects of conditions speaker origin, word origin, beginners and advanced learners L1 vs. L2, and beginners and advanced learners in L2. Tests with gender as factor showed no differences between male and female speakers. Therefore, we do not differentiate gender in the presentation of the results.

3. Results

3.1. German speakers vs. French speakers

The vowel productions of GS in GW show a tendency for more strained phonatory behavior in comparison to the productions of FS in FW. This behavior is illustrated in Figure 3 for the parameter GOG. The parameter value provides information about the degree of glottal opening. The significantly smaller mean value of the parameter for GS – compared to that of the FS – indicates this phonatory behavior. Thus, FS show different adduction behavior expressed by a more incomplete glottal closure during vowel production.

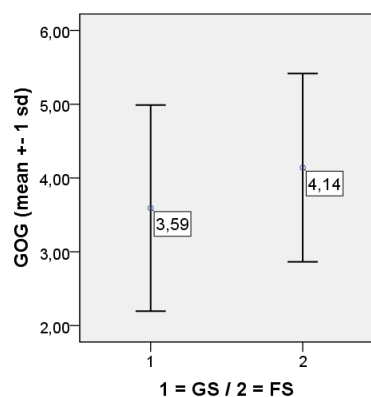


Figure 3: Glottal opening gradient (GOG); $p < 0.001$.

3.2. Vowel productions L1 vs. L2

3.2.1. German speakers

The vowel productions of German words (GW) produced by GS are significantly different in comparison to their French words (FW). This different phonatory behavior is instrumentally shown in Figure 4 by higher parameter values of the two parameters OQG (OQG stands for open quotient and provides information about adduction behavior of the vocal folds) and GOG when FW are produced. Higher parameter values point to less adduction of the vocal folds and a more incomplete glottal closure during production.

3.2.2. French speakers

FS show less adduction behavior when they produce vowels in GW. This tendency is shown in Figure 5. The means of the parameter IC reflect the bandwidth of the first formant. A larger bandwidth caused by less adduction behavior of the vocal folds leads to higher values of this parameter. An increasingly higher loss of acoustic energy in the subglottal cavity can be assumed.

It is interesting to note that for GS differences in phonatory behavior during the production of GW and FW are shown by significantly different values of the spectral gradient parameters (OQG, GOG). For FS however, these differences are shown by values of the bandwidth parameter IC. It might be the case that GS (intuitively) focus more on timing aspects of the glottal cycle, whereas FS rather focus on the degree of glottal opening. Note that only these three parameters are relevant for the above-mentioned comparison.

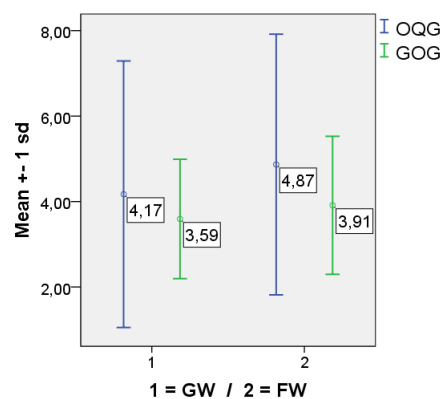


Figure 4: German speakers (GS): Open quotient gradient (OQG) and Glottal opening gradient (GOG); $p < 0.01$.

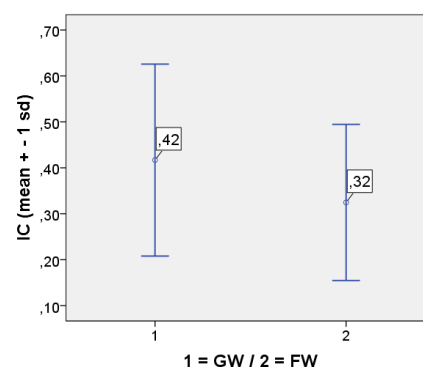


Figure 5: French speakers (FS): Incompleteness of closure (IC); $p < 0.001$.

3.3. Vowel productions of beginners and advanced learners L1 vs. L2

In a third step, we look at potential differences between BEG and ADV of the two speaker groups concerning their vowel productions in L2 words compared to those in L1 words.

3.3.1. German speakers

BEG show changes in phonatory behavior when they produce vowels in L2 words. These changes are demonstrated by less adduction of the vocal folds. In Figure 6 an example of this behavior is given by means of the OQG parameter values (see parameter means on the left). The significantly smaller adduction – in comparison to L1 productions – can be pointed out by a higher mean value of this parameter. Further, the obtained significantly different parameter values ($p < 0.001$, $p < 0.01$) of the parameters GOG, SKG (skewness of the glottal pulse), RCG (rate of closure), and T4G (fourth formant) underline these findings. Additionally, ADV show the same tendency, but, there are no significant differences between L1 and L2 vowel productions (see parameter means of OQG on the right).

3.3.2. French speakers

Values of the two parameters GOG and IC for BEG and ADV show significant differences between L2 and L1 vowel productions. BEG display a more incomplete glottal closure expressed by a higher GOG value (see Figure 7 on the left) when they produce vowels in L2 words, whereas ADV show

the opposite tendency. Their production behavior is orientated on the behavior of the GS with more strained vocal fold vibrations. The smaller mean value of the GOG parameter indicates this behavior (see Figure 7 on the right).

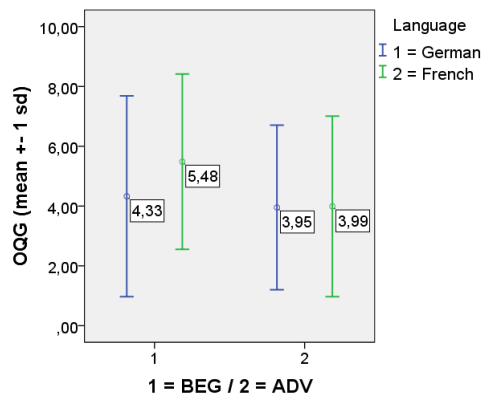


Figure 6: German speakers (GS): *Open quotient gradient (OQG)*; BEG = $p < 0.01$; ADV = not significant.

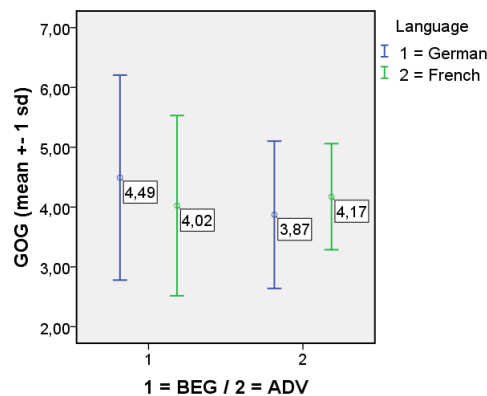


Figure 7: French speakers (FS): *Glottal opening gradient (GOG)*; BEG = $p < 0.01$; ADV = $p < 0.05$.

3.4. Vowel productions of beginners and advanced learners in L2

It is apparent from Figures 6 and 7, respectively, that ADV of the two speaker groups (GS and FS) show a tendency towards more strained phonatory behavior in vowel production of L2 words in comparison to BEG (see OQG and GOG values). Apparently, it can be pointed out that German ADV seem to maintain their production behavior in L2 (see Figure 3 and 6), whereas French ADV seem to change theirs. A more strained position of the vocal folds during phonation can be assumed (see Figure 3 and 7).

4. Discussion

In this study, measurements from the acoustic signal were used to characterize phonatory behavior of German (GS) and French (FS) speakers in native speech, and subsequently, to evaluate changes of this behavior of the two groups in non-native speech.

It can be firstly pointed out that the phonatory behavior of GS is characterized – in comparison to the behavior of FS – by more strained adduction of the vocal folds. For this speaker group a higher subglottal pressure, a stronger contraction of the laryngeal muscles as well as a shorter duration of closing movements can be concluded physiologically. This behavior is

instrumentally shown by smaller values of the acoustic parameter GOG (glottal opening gradient). However, FS show – in comparison to GS – a more incomplete adduction expressed by higher values of this parameter when they produce vowels in their L1 (see Figure 3). Furthermore, the comparison of productions in L2 words with productions in L1 words show that GS change their phonatory behavior in L2 by adapting their adduction behavior to that of FS, whereas FS do not show this tendency. They even produce GW with less vocal fold adduction than in their L1 (see Figure 4 and 5).

A comparison of phonatory behavior of BEG vs. ADV of the two speaker groups suggests the following: German BEG are already orientated on production characteristics of L2 by changing their phonatory behavior significantly in L2. German ADV only show a tendency in this direction (see Figure 6). French BEG retain their production behavior by showing – in comparison to their L1 productions – less vocal fold adduction in L2 (see Figure 7). French ADV however show more strained adduction so that their phonatory behavior illustrates the opposite. Finally, ADV of the two speaker groups (GS and FS) generally show more strained behavior in L2 productions than BEG (see Figures 3, 6 and 7). This suggests that FS seem to adapt their phonatory behavior in L2 whereas GS seem to maintain it.

The two speaker groups show different laryngeal phonatory settings (also so called articulatory settings) when they produce vowels in L1 in comparison to their L2 production [32]. This is in line with claims regarding different voice quality patterns in cross-linguistic comparisons (e.g. [21]). FS and GS seem to apply different postural language-specific configurations. These configurations have to be deployed by the laryngeal articulators when vowel, and expectedly consonant productions in the two different languages have to be produced [33, 34].

5. Conclusions

The results of the study provide evidence that the German and French speakers apply different laryngeal phonatory settings and that they alter their settings in L2 *differently*. Furthermore, our findings reported here are somewhat different from results reported in [22, 23] where LTAS measures showed parallel tendencies for different speaker groups in L2. However, the use of a different and more physiologically orientated approach to measure phonatory behavior and phonatory quality might be crucial for these differences. Using this approach a closer examination of phonatory behavior in L1 and L2 productions seems to be possible. Finally, a perceptual evaluation of voice quality of the speech material using a rating scale should be considered as a next research step. Subsequently, it would be interesting to investigate whether a correlation exists between the results of the rating experiment and the tendencies of phonatory behavior characterized in the present study. In doing so, a contribution to the yet not well understood relation between acoustic-phonetic similarity or dissimilarity and perceptual similarity or dissimilarity of L1/L2 would be given.

6. Acknowledgements

This research was supported by DFG and ANR for the project “IFCASL” (Individualized Feedback for Computer-Assisted Spoken Language Learning (www.ifcasl.org)).

7. References

- [1] C. T. Best, "A direct realist view of cross-language speech perception," in W. Strange [Ed], *Cross-language studies of speech perception: A historical review*, York: Timonium, 1995, pp. 171-206.
- [2] J. E. Flege, M. J. Munro, and R. A. Fox, "Auditory and categorical effects on cross-language vowel perception," *Journal of the Acoustical Society of America*, vol. 95, pp. 3623-3641, 1994.
- [3] J. Kingston, "Learning foreign vowels," *Language and Speech*, vol. 46, pp. 295-349, 2003.
- [4] A. R. Bradlow, R. Akahane-Yamada, D. B. Pisoni, and Y. I. Tohkura, "Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production," *Perception & Psychophysics*, vol. 61, pp. 977-985, 1999.
- [5] A. R. Bradlow, D. B. Pisoni, R. Akahane-Yamada, and Y. I. Tohkura, "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," *Journal of the Acoustical Society of America*, vol. 101, pp. 2299-2310, 1997.
- [6] J. Anderson-Hsieh, R. Johnson, and K. Koehler, "The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure," *Language Learning*, vol. 42, pp. 529-555, 1992.
- [7] M. Jilka, "Different manifestations and perceptions of foreign accent in intonation," in J. Trouvain, U. Gut [Eds], *Non-Native Prosody - Phonetic Description and Teaching Practice*, ed Berlin: Mouton de Gruyter, 2007, pp. 77-96.
- [8] P. Trofimovich and W. Baker, "Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech," *Studies in Second Language Acquisition*, vol. 28, pp. 1-30, 2006.
- [9] M. J. Munro, T. M. Derwing, and C. S. Burgess, "Detection of nonnative speaker status from content-masked speech," *Speech Communication*, vol. 52, pp. 626-637, 2010.
- [10] W. Barry, B. Andreeva, and I. Steiner, "The phonetic exponency of phrasal accentuation in French and German," in *Proceedings of Interspeech 2007*, Antwerp, Belgium, 2007, pp. 1010-1013.
- [11] I. Mennen, F. Schaeffler, and C. Dickie, "Second Language Acquisition of Pitch Range in German Learners of English," *Studies in Second Language Acquisition*, vol. 36, pp. 303-329, 2014.
- [12] P. Keating and G. Kuo, "Comparison of speaking fundamental frequency in English and Mandarin," *The Journal of the Acoustical Society of America*, vol. 132, pp. 1050-60, Aug 2012.
- [13] J. Caspers and O. Kepinska, "The influence of word-level prosodic structure of the mother tongue on production of word stress in Dutch as a second language," in *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, 2011, pp. 420-423.
- [14] K. Aoyama and S. G. Guion, "Prosody in second language acquisition: Acoustic analyses of duration and F0 range," in O.-S. Bohn and M. J. Munro [Eds], *The role of language experience in second-language speech learning - In honor of James Emil Flege*, ed Amsterdam: John Benjamins, 2007, pp. 281-297.
- [15] B. Andreeva, G. Dementko, B. Möbius, F. Zimmerer, J. Jügler, and M. Oleskowicz-Popiel, "Differences of Pitch Profiles in Germanic and Slavic Languages," in *Proceedings of Interspeech 2014*, Singapore, 2014.
- [16] B. Andreeva, B. Möbius, G. Dementko, F. Zimmerer, and J. Jügler, "Linguistic measures of pitch range in Slavic and Germanic languages," in *Proceedings of Interspeech 2015*, Dresden, 2015, pp. 968-972.
- [17] R. van Bezooijen, "Sociocultural aspects of pitch differences between Japanese and Dutch women," *Language and Speech*, vol. 38, pp. 253-265, 1995.
- [18] R. Hincks and J. Edlund, "Promoting increased pitch variation in oral presentations with transient visual feedback," *Language Learning & Technology*, vol. 13, pp. 32-50, 2009.
- [19] M. G. Busà and M. Urbani, "A cross linguistic analysis of pitch range in English L1 and L2," in *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, 2011, pp. 380-383.
- [20] R. Ullakonoja, "Comparison of pitch range in Finnish (L1) and Russian (L2)," in *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI)*, Saarbrücken, 2007, pp. 1701-1704.
- [21] J. H. Esling, "Crosslinguistic aspects of voice quality," in R. D. Kent and M. J. Ball [Eds], *Voice quality measurement*, ed San Diego, CA: Singular Publishing Group, 2000, pp. 37-48.
- [22] A. Ruß, *Stimmqualität zwischen Mutter- und Fremdsprache - Eine Produktionsstudie zum deutsch-italienischen Sprachenpaar*: AV Akademikerverlag, 2015.
- [23] A. P. P. F. Engelbert, "Cross-linguistic effects on Voice Quality: A study on Brazilians' production of Portuguese and English. International Symposium on the Acquisition of Second Language Speech," *Concordia Working Papers in Applied Linguistics (COPAL)*, vol. 5, pp. 157-170, 2014.
- [24] W. Wokurek and M. Pützer, "Automated corpus based spectral measurement of voice quality parameters," in *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS XV)*, Barcelona, Spain, 2003, pp. 2173-2176.
- [25] M. Pützer and W. Wokurek, "Phonatory characteristics and voice quality evaluation of Laryngeal Dystonia before and after Botulinum toxin treatment. A case study," in *Proceedings of the 10th International Conference on Voice Physiology and Biomechanics (ICVPB)*, Viña del Mar, Chile, 2016.
- [26] M. Pützer, W. J. Barry, and J. R. Moringlane, "Effect of deep brain stimulation on different speech subsystems in patients with multiple sclerosis," *J Voice*, vol. 21, pp. 741-53, Nov 2007.
- [27] M. Pützer and W. J. Barry, "Instrumental dimensioning of normal and pathological phonation using acoustic measurements," *Clin Linguist Phon*, vol. 22, pp. 407-20, Jun 2008.
- [28] J. Trouvain, Y. Laprie, B. Möbius, B. Andreeva, A. Bonneau, V. Colotte, et al., "Designing a bilingual speech corpus for French and German language learners," in *Proceedings of the Corpus et Outils en Linguistique, Langues et Parole: Statuts, Usages et Mélanges*, Strasbourg, 2013, pp. 32-34.
- [29] C. Fauth, A. Bonneau, F. Zimmerer, J. Trouvain, B. Andreeva, V. Colotte, et al., "Designing a bilingual speech corpus for French and German language learners: a two-step process," in *Proceedings of the LREC 2014*, Reykjavik, Iceland, 2014, pp. 1477-1482.
- [30] P. Boersma and D. Weenink, *PRAAT: Doing phonetics by computer*, Version 5.4 ed., 2014.
- [31] K. N. Stevens and H. M. Hanson, "Classification of glottal vibration from acoustic measurements," in O. Fujimura and M. Hirano [Eds], *Vocal fold physiology: voice quality control*, ed San Diego: Singular Publishing Group, 1995, pp. 147-170.
- [32] J. Laver, *The Phonetic Description of Voice Quality*. Cambridge, England: Cambridge University Press, 1980.
- [33] V. Ramanarayanan, L. Goldstein, D. Byrd, and S. Narayanan, "An MRI study of articulatory settings of L1 and L2 speakers of American English," in *Proceedings of the 9th International Seminar on Speech Production*, Montreal, Canada, 2011.
- [34] I. Mennen, J. M. Scobbie, E. de Leeuw, S. Schaeffler, and F. Schaeffler, "Measuring language-specific phonetic settings," *Second Language Research*, vol. 26, pp. 13-41, 2010.